

醫 "Ae151", "良醫 "Ae151", "賢醫 "Ae151", "名醫 "Ae151", "神醫 "Ae151", "太醫
" "Ae151", "御醫 "Ae151", "法醫 "Ae151", "仵作 "Ae151", "世醫 "Ae151", "儒醫"
"Ae151", "庸醫 "Ae151", "西醫 "Ae151", "牙醫 "Ae151", "獸醫 "Ae151", "軍醫"
"Ae151", "廠醫 "Ae151", "校醫 "Ae151", "赤腳醫生 "Ae151", "主任醫師 "
"Ae151", "副主任醫師 "Ae151", "主治醫師 "Ae151", "住院醫師 "Ae151", "醫士"

肆、語料庫與計算語言學

利用語料庫與統計是近年來計算語言學研究的主要趨勢。無論對語音辨識，語法剖析，歧義的解決，機器翻譯，與詞彙知識的自動取得在在都需要大型語料庫。語音辨識是將語音訊號轉變成文字，基本上可以分成前處理與後處理。前處理是從聲波的物理性質猜測最有可能的母音與子音組合。而後處理則從最有可能的音中選出最有可能的詞，無論是語音辨識的前處理或後處理或解決詞類歧義目前最常用的統計理論是隱式馬可夫模型(Hidden Markov Model 簡稱 HMM)。而解決結構歧義則常利用語法樹庫計算某一詞出現在某一種結構的機率，例如 John saw the man with a telescope.其中的介詞組 with a telescope 可以修飾名詞組 the man(約翰看見一個帶望遠鏡的人)也可以修飾動詞 saw(約翰用望遠鏡看見一個人)，造成歧義的現象。英文這種所謂 PP attachment 結構歧義的問題跟詞彙的語義與語用有關，過去計算語言學家嘗試用規則來處理效果不好，目前改以語法樹庫計算介詞組內的名詞分別跟受詞與動詞的相關性機率，從而預測介詞組究竟修飾受詞或動詞。統計演算法不需大量的人力來撰寫語言規則或編纂語言知識，可以從大型語料庫中直接抽取諸如同義詞，反義詞，搭配語等語言知識。統計方式也可以自動抽取部分中文詞彙。